APPENDIX H

Input: 1. A continuous dataset X1

2. A categorical dataset X2

Output: X1 untouched, X2 may get smaller by dropping some variables

Parameter: CV, a threshold for Crammer's V value

Process:

1. Bin each continuous variable into a number of categories (if not already performed).

2. For each of the categorical variable x2, compute the Crammer's V value between x2 and each binned continuous one.

If the Cramer's V > the threshold CV,

Drop the categorical

End If